

AD-772 501

ON BALANCE AND INITIALIZATION

Michael Ghil

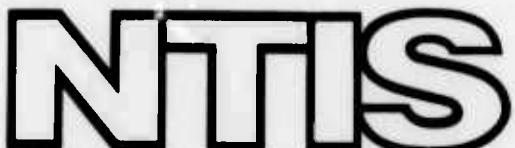
New York University

Prepared for:

Advanced Research Projects Agency
Army Research Office-Durham

December 1973

DISTRIBUTED BY:



National Technical Information Service
U. S. DEPARTMENT OF COMMERCE
5285 Port Royal Road, Springfield Va. 22151

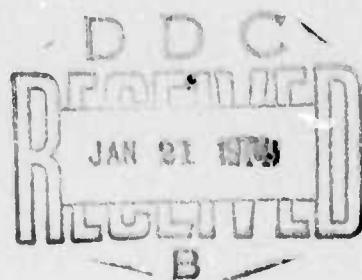
IMM 400
December 1973

Courant Institute of
Mathematical Sciences

AD772501

On Balance and Initialization

Michael Ghil



Sponsored by Advanced Research Projects Agency,
ARPA Order No. 2069

Approved for public release; distribution unlimited.



New York University

Directed by
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. Department of Commerce
Springfield, VA 22161

ARPA Order No.: 2069

IMM 400
December 1973

New York University
Courant Institute of Mathematical Sciences

ON BALANCE AND INITIALIZATION

Michael Ghil

This research was supported by the Advanced Research Projects Agency under Grant No. DA-ARO-D-31-124-72-G113.

Views and conclusions contained in this study should not be interpreted as representing the official opinion or policy of the Courant Institute of Mathematical Sciences, or of New York University, or of ARPA.

Approved for public release; distribution unlimited.

Acknowledgements

This work was started when the author was at the Goddard Institute for Space Studies as a staff member of Computer Sciences Corporation. He wishes to thank Dr. K. Grossman, then of CSC, for suggesting the problem and Dr. M. Halem of GISS for continuous encouragement. It is a pleasure to acknowledge valuable discussions with Prof. J. G. Charney, Dr. S. Schneider and Dr. T. Gal-Chen on the meteorological background and history of the problem. Discussions with Prof. E. Isaacson and Prof. P. D. Lax concerning the mathematical aspects of the problem were very helpful. Prof. E. Isaacson also carefully read the manuscript and this is a welcome opportunity to thank him once more for his varied and pertinent comments.

Abstract

The initialization problem is defined as the problem of obtaining the initial data that are required in order to solve a well posed initial-and-boundary-value problem for the equations of large scale dynamical meteorology. In our treatment of this problem complete knowledge at a given instant of the atmospheric temperature and surface-pressure fields, as well as of their time derivatives, is assumed to be at our disposal; and the problem of computing the velocity field at that instant from this knowledge is studied.

The methods used are based on the exact mathematical treatment of the differential equations of atmospheric motion, rather than on perturbation-type or numerical methods. A simple solution of the initialization problem thus formulated is exhibited in a particular case. It is shown, however, that in general the problem is indeterminate and additional information is required to uniquely find the velocity field. Four-dimensional data assimilation is also discussed.

1. Introduction

Shortly after the development of computers made the numerical integration of the primitive equations[†] over a global grid feasible, it became apparent that one of the main problems in such an integration was to provide the initial data needed to start it. Indeed, the primitive equations, in x,y,z-coordinates, say, form a totally hyperbolic system, when viscosity and heat conduction are neglected.* A well-posed problem for this system would be to specify throughout the global atmosphere the velocity components u , v , w , and the temperature T at a given time $t = t_0$, as well as certain conditions, into the details of which we will not enter, at the upper and at the lower boundary of the atmosphere. The complete and sufficiently accurate specification of such initial conditions became known as the initialization problem.

Thus the problem falls into two parts: (i) the completeness, and (ii) the accuracy of the initial data. In discussing the different aspects of this problem, we will try to distinguish between the physical behavior of the atmosphere, the mathematical (differential) models describing this behavior, and the numerical (difference) models used to approximate the mathematical model. The behavior of a numerical model is often related directly to

[†] The Eulerian equations of motion with the vertical momentum equation replaced by the hydrostatic assumption.

* Assuming the surface pressure is known, the pressure p can be obtained directly by integrating the hydrostatic equation, which we exclude from this discussion.

and compared with that of the atmosphere, the intermediate level of approximation provided by the mathematical model from which it was derived being omitted from the discussion or the behavior of the mathematical model being tacitly implied to be the same as either that of the atmosphere, or as that of the numerical model. We shall briefly mention first the questions related to the accuracy problem.

When finite-difference approximations to the primitive equations started to be used for meteorological research and weather prediction, it was noted that the numerical models* simulated not only the large-scale motions of the atmosphere, in which one was interested, but also faster phenomena of smaller amplitude which were identified as the numerical counterpart of inertia-gravity waves, and which are adequately represented for instance by solutions of the linearized shallow-fluid equations (Charney, 1955). These waves were undesirable and their cause was traced to the lack of geostrophic adjustment in the initial data used. However, it was soon realized that initial conditions, satisfying any hitherto used balance equation, would not prevent the existence and propagation of inertia-gravity waves in a primitive-equation model (Nitta and Hovermale, 1969, Morel et al., 1971). The mathematical reason for this is the fact

* In this context a numerical model of the atmosphere is essentially a finite-difference approximation to the equations of motion. It includes in general also some terms which represent the direct numerical parametrization of certain physical processes, but we will not dwell upon this aspect of the models, considering them as difference schemes for the integration of the equations of motion, i.e., of the mathematical model.

that the forms of the balance equation* known from previous quasi-geostrophic and quasi-nondivergent models do not represent an instantaneous compatibility condition between wind and pressure field, consistent with the primitive equations describing the motion. Actually such a balance equation, in its general form or in any of its simplified forms, represents only an approximation to such a condition, obtained by neglecting certain terms in the equations from which it is derived. Indeed, the atmosphere itself is only approximately in geostrophic balance, rather than exactly in this state.

It appeared though that the amplitude of the fast waves in the numerical model far exceeded that of the inertia-gravity waves in the atmosphere. These large amplitudes could reasonably be attributed to an excessive imbalance in the initial conditions, due to the use of data incorrectly measured or interpolated. Since this situation could not be remedied by traditional synoptic balancing, Nitta and Hovemale (1969) proposed a dynamical balancing. The procedure consisted essentially in running the model successively for short periods forward and backward in time for a number of cycles. It was expected that the model would simulate geostrophic adjustment, i.e., disperse and dissipate the fast, short-scale waves. Positive improvement was obtained with this method.

* For the most general form of the balance equation, as well as for some of the more special forms it takes when certain simplifying assumptions are made, see Haltiner (1971).

The reduction of the exaggerated amplitude of small-scale motions produced by initial errors was known to be necessary in order to prevent them from influencing too strongly in too short a time span the large-scale motions represented by the numerical model. The relationship between the size of initial errors and the predictability range* has been first investigated systematically by Charney et al. (1966). For the presently available accuracy in measurement, the predictability range of the existing numerical models of the atmosphere is believed to be about two weeks, assuming that complete information with the given accuracy were available.

This brings us back to the completeness problem. In fact, there is at present no meteorological data gathering system capable of supplying u , v , w , T over a uniform grid, like the ones used by general circulation models (GCMs), at a given time ("synoptically"). However, it seems reasonable to expect from the progress of satellite technology at least good date coverage for temperatures in the near future. Therefore it has been suggested by Charney et al. (1969) that the wind field (u, v, w) , speaking vaguely for the moment, could be inferred from the temperature field T and some of its past history.

The rationale for this proposal was twofold. First, the experience with barotropic models, i.e., with systems of partial

* The predictability range is defined as the time span after which the errors in the calculated state exceed a certain pre-assigned threshold, usually related to the difference between two "randomly chosen" states of the atmosphere.

differential equations in which the barotropic approximation was made, and in which the wind field could be obtained from the instantaneous geopotential field by solving a time-independent balance equation. Numerical integrations of these equations were rather successful, approximating well certain aspects of the behavior of the atmosphere. Second, the well-known mathematical fact that, in a linear hyperbolic system of n first-order partial differential equations with constant coefficients, $(n-1)$ dependent variables can be eliminated, thereby reducing the system to a single equation for the n -th variable. Then the initial conditions on the $(n-1)$ eliminated variables lead to conditions for the time derivatives up to order $(n-1)$ of the n -th variable (e.g., Courant and Hilbert, 1962, p. 14 ff.). In particular, this is the case in linear acoustics, where the velocity components can be eliminated, to yield a single equation for the pressure. The discussion of these considerations will be briefly taken up in Section 4.

The concrete procedure proposed by Charney et al. (1969) was to start the numerical integration of the equations with the "correct" temperature data and with velocity data approximated by some other method, then, at given time intervals, to replace the calculated values of T by the "correct" ones (i.e., to "update" T). In fact, however, no "correct" measured (satellite) temperature data were available. Therefore the experiments were made to test the procedure by relying on the concept of a "control run", i.e., of a numerical integration of the equations performed for a certain length of time, the results of which are referred to for

purposes of comparison as the "true history" of the atmosphere over that time span.

First the control run, with certain initial values of u , v , w , T , is performed and the values of T corresponding to this run are taken as "correct". Then another integration is started, with the same initial temperature data, but different initial velocity data, and T is "updated" with values taken from the "correct" temperatures of the control run. The reason for updating the temperature is to reduce the error in the velocity, caused by using inaccurate initial velocity data.

It was indeed observed that the root-mean-square errors in the velocity field decreased to a certain non-zero asymptotic value. However, in the experiments reported by Charney et al. (1969), as well as in similar experiments performed later (e.g. Williamson and Kasahara, 1971) this value was reached only after a time comparable to the predictability range of the numerical models used (10 days to 3 weeks) and this asymptotic rms error was non-negligible, although considerably smaller than the rms value of the initial error, and a fortiori much smaller than the rms error in the wind field after a similar period in a run without updating. These facts seem to limit the value of the updating approach to the completeness problem in spite of the partial success of the method.

More recently (Morel et al., 1971, Mesinger, 1972) the method of iterative dynamic adjustment, suggested by Nitta and Hovermale (1969) in dealing with the geostrophic balance problem, was applied to the data acquisition problem. This was done in

order to circumvent the questions arising from achieving proper initialization only at the end of the predictability time. Good results were reported with proper choice of finite-difference scheme, period of the back-and-forth iteration, etc. There seems however to be a marked distinction between the results using data from a control run of the same model and those when using real data or data from a different model. The results when using "extraneous" data for iterative updating are much worse and rather discouraging (Morel et al., 1971, Mesinger, 1972).

It may be worthwhile to stress that the forward-updating and iterative techniques discussed before were proposed when it became clear that traditional objective analysis methods, which had been fairly successful for the "filtered" (quasi-geostrophic and quasi-nondivergent) models, produced "initialization shocks" and large-amplitude short waves in primitive equation models. However, the consensus seems to be that up to now these asynchronous techniques are much more computer-time consuming and do not seem to have performed convincingly better than the more sophisticated ones among the objective-analysis methods (including least-squares and weighted-average fits of observed data, variational techniques, weighted consideration of prognostic computed data, permanence and climatology data, as well as use of geostrophic and balanced data). Therefore it is still the synchronous objective analysis methods which are operational in daily forecasting (Haltiner, 1971) and it is one of the purposes of the search for better initialization methods to develop tools for improved and extended operational forecasts.

In view of the difficulties encountered by the non-synoptic techniques, as well as by the traditional synoptic techniques, we shall try to pursue a different avenue of attack on the initialization problem. As we saw, the two aspects of the problem, completeness and accuracy, are strongly interrelated and any solution has to take care of both. This brings us back to the discussion of well-posed problems for the system of equations dealt with, viz., the primitive equations (where we assume the surface pressure to be known and hence ignore the hydrostatic equation), i.e., to the question of the side conditions under which the existence of a unique solution of this system, continuously depending on the data, is guaranteed.

Given the hyperbolic character of the equations and their non-linearity it seems mathematically very implausible that any other problem, but the proper initial- and boundary-value problem which one would wish to solve, would be well posed. In particular, if in any sense the temperature field determines the wind field and if a state of quasi-geostrophic equilibrium, meaning a state of minimum amplitude gravity-inertia waves, exists, these two facts ought to reflect an instantaneous property of the primitive equations, rather than an integral property involving a finite time span. This point will be further discussed in Section 4. Such an instantaneous property would have to be expressed mathematically as one or more equations involving the velocity components and their space derivatives, but not their time derivatives, and some variables of state (such as tempera-

ture and pressure), which are assumed to be known together with any number of their space and time derivatives.

The presence of as many of the time derivatives of state variables as necessary would generalize the usual forms of the balance equation, which are completely free of time derivatives, and these derivatives would be all that is needed to express the thermodynamic history of the system. These time derivatives could be computed from sufficiently many discrete measurements by using adequate numerical techniques, which probably would have to include certain space-averaging procedures.

If the number of independent equations thus obtained from the primitive equations by differentiation and elimination, together with proper side conditions, suffices to determine the wind field at a given instant, then we will have solved the completeness-of-data problem, assuming adequate knowledge of the thermodynamic history. If not, the equation or equations obtained will still express a compatibility condition, equivalent in the general baroclinic case to the balance equation. This condition will then be the only limitation imposed by the equations themselves on the initial data and thus the only "dynamic" check on the accuracy of these data.

It is hoped that the following analysis will clarify this point of view and indicate some of its possibilities and limitations.

2. Compatibility conditions for the shallow-fluid equations

As an example of the ideas involved in the approach we want to discuss, we consider the shallow-fluid equations for a rotating Cartesian x, y -coordinate system

$$(1a) \quad u_t + uu_x + vu_y + \phi_x - fv = 0 ,$$

$$(1b) \quad v_t + uv_x + vv_y + \phi_y + fu = 0 ,$$

$$(1c) \quad \phi_t + u\phi_x + v\phi_y + \phi(u_x + v_y) = 0 ,$$

$$\phi = gh .$$

Here u, v are the velocity components in the x, y directions respectively, h is the height of the free surface, g the acceleration of gravity and f the Coriolis parameter. For the sake of simplicity we assume f to be a constant.

First we will study the linearization of this system around a state of rest. With an obvious change of notation we then have

$$(2a) \quad u_t + \phi_x - fv = 0 ,$$

$$(2b) \quad v_t + \phi_y + fu = 0 ,$$

$$(2c) \quad \phi_t + \Phi(u_x + v_y) = 0 ,$$

$$(2d) \quad \Phi = \text{const.}$$

Here Φ is the equilibrium value of the geopotential gh of the free surface, and ϕ stands now for $gh - \Phi$, the deviation of the geopotential from Φ .

Differentiating (2c) with respect to t , (2a) and (2b) with respect to x and y , we obtain

$$(3a) \quad u_{tx} + \phi_{xx} - fv_x = 0 ,$$

$$(3b) \quad v_{ty} + \phi_{yy} + fu_y = 0 ,$$

$$(3c) \quad \phi_{tt} + \Phi(u_{xt} + v_{yt}) = 0 .$$

Substituting u_{xt} and v_{yt} from (3a), (3b) into (3c) yields

$$(4) \quad \Phi f(v_x - u_y) - \Phi(\phi_{xx} + \phi_{yy}) + \phi_{tt} = 0 .$$

Using our standing assumptions that ϕ and as many of its derivatives as necessary are known, we obtain from (4) and (2c) the following system of two first-order partial differential equations for u , v

$$(5) \quad \begin{aligned} u_x + v_y &= -\phi_t/\Phi , \\ u_y - v_x &= (\phi_{tt} - \Phi \Delta \phi)/\Phi f , \end{aligned}$$

where Δ is the two-dimensional Laplacian operator,

$$\Delta \phi = \phi_{xx} + \phi_{yy} .$$

Thus we see that in this very simple case it is possible to eliminate the time derivatives u_t , v_t and obtain the instantaneous compatibility conditions (5).

System (5) for the functions u and v is elliptic, i.e., it has no real characteristic. In fact it is just a set of inhomogeneous Cauchy-Riemann equations for the functions v , u , and by

cross-differentiation these equations lead to a Poisson equation for either u or v . A well-posed problem for (5) would be the Dirichlet problem. This would mean prescribing u say on a closed contour ∂D and v at some point on the contour or in its interior, D .

It is well known (Morel et al., 1971, Williamson and Dickinson, 1972) that the system (2) has three independent plane-wave solutions, one corresponding to slow Rossby waves, the other two to inertia-gravity waves propagating in opposite directions. In the case at hand, where the unperturbed velocity is zero, the Rossby mode is stationary and the inertia-gravity waves have phase-velocity

$$c = \pm (k_\Phi^2 + f^2)^{1/2} / k, \quad k^2 = k_1^2 + k_2^2,$$

$\underline{k} = (k_1, k_2)$ being the wave vector. Any solution of (2) can be represented by a series expansion in these plane waves.

Let a quantity X be decomposed as

$$X = \bar{X} + X',$$

where $(\bar{\cdot})$ is a time average and hence \bar{X} , being stationary, is representable in terms of the Rossby mode, whereas X' stands for the inertia-gravity-mode part of the representation. Then

$$(2') \quad f\bar{u} = -\bar{\phi}_y, \quad f\bar{v} = \bar{\phi}_x$$

is a geostrophically balanced stationary solution, satisfying eqs. (2) as well as system (5) with $\partial/\partial t = 0$.

The most general stationary solution of (2) will satisfy

$$(6a) \quad u_x + v_y = 0 ,$$

$$(6b) \quad v_x - u_y = \Delta \chi ,$$

where $\chi = \phi/f$. The solution of (6a) is

$$(7) \quad u = -\psi_y , \quad v = \psi_x ,$$

with ψ an arbitrary, twice continuously differentiable function.

By (6b) ψ has to satisfy the equation

$$(8) \quad \Delta(\psi - \chi) = 0 .$$

Hence if

$$\psi = \chi \quad \text{on } \partial D ,$$

then

$$\psi = \chi \quad \text{in } D ,$$

where D is some domain and ∂D its boundary. Similarly, if

$$\partial_n \psi = \partial_n \chi \quad \text{on } \partial D ,$$

where

$$\partial_n = n \cdot \nabla$$

and n is the unit normal to ∂D , ∇ the gradient, then

$$(*) \quad \psi = \chi + \text{const.} \quad \text{in } D .$$

But we have from (7) that

$$\partial_n \psi = n \cdot (v, -u) .$$

Therefore the boundary condition

$$(9) \quad n \cdot (v, -u) = \partial_n \chi \quad \text{on } \partial D ,$$

and in particular

$$(10) \quad u = -\chi_x , \quad v = \chi_y \quad \text{on } \partial D ,$$

implies, by (*), that

$$u = -\chi_x , \quad v = \chi_y \quad \text{in } D .$$

Hence we can conclude by a continuity argument that the only case in which (9) can hold is that (10) be satisfied. An obvious special instance of the Neumann problem above is that in which the boundary ∂D is a level line of the height field,

$$\partial D: \quad \chi = \text{const.}, \quad h = \text{const.},$$

in which case

$$u = -\chi_y , \quad v = \chi_x \quad \text{in } D ,$$

provided ∂D is also a streamline of the velocity field (u, v) and

$$u^2 + v^2 = (\partial_n \chi)^2 \quad \text{on } \partial D .$$

Generally speaking we can state that the time-averaged solutions $(\bar{u}, \bar{v}, \bar{\phi})$ of (2) will be geostrophic, i.e., satisfy (2'), in a domain D if and only if they are "geostrophic" on ∂D . Moreover, because of the maximum principle for the Laplace equation, deviations ϵ from geostrophicity in the domain D ,

$$\bar{e} = \max \{ |fu + \bar{\phi}_y|, |fv - \bar{\phi}_x| \} ,$$

will be bounded by the values of \bar{e} on ∂D .

Thus we see that in the simple flow described by (2) we can solve the completeness problem in initialization by determining u , v from the compatibility conditions or "generalized balance equations" (5), given ϕ , ϕ_t and ϕ_{tt} .

Moreover, a nearly geostrophic wind field can be obtained from a time-dependent geopotential ϕ in the following manner. First expand the measured ϕ into plane waves, as indicated before. Then suppress the inertia-gravity waves, i.e., set their coefficients to zero, and call the new function thus obtained ϕ' . Clearly $\phi'_t = 0$, $\phi'_{tt} = 0$, so that instead of (5) we will have for u , v a system of the form (6), with $\chi = \phi'/f$. On the boundary, however, the departure e from geostrophicity,

$$e = \max \{ |fu + \phi'_y|, |fv - \phi'_x| \} ,$$

will in general not be zero. But we have shown that for system (6) the error e in the interior of D will be bounded by the value of e on the boundary ∂D . Hence it suffices to adequately modify u , v on ∂D and/or ϕ' on ∂D (and near ∂D , to preserve continuity) in order to keep e throughout D within prescribed limits. This procedure is in a sense comparable to that used in the analysis of Williamson and Dickinson (1972).

We are now in a position to return to the more general case (1) and ask whether the complete analysis performed for (2) generalizes. Unfortunately the answer seems to be that it does not.

To obtain compatibility conditions similar to (5) for the system (1) we would have again to eliminate u_t , v_t from it by differentiating (1c) with respect to t . Before doing this let us simplify (1c) by introducing

$$\varphi = \log \phi .$$

Then φ and its derivatives will be known, since ϕ and its derivatives are known by our standing assumption. Dividing now (1c) by $\phi > 0$, it becomes

$$(11) \quad u_x + v_y + \varphi_x u + \varphi_y v + \varphi_t = 0 ,$$

where φ_x , φ_y , φ_t are known. Differentiating (11) with respect to t , (1a) and (1b) with respect to x and y , we obtain

$$(12a) \quad u_{tx} + uu_{xx} + vu_{xy} + u_x^2 + u_y v_x - fv_x + \varphi_{xx} = 0 ,$$

$$(12b) \quad v_{ty} + uv_{xy} + vv_{yy} + u_y v_x + v_y^2 + fu_y + \varphi_{yy} = 0 ,$$

$$(12c) \quad u_{xy} + v_{yt} + \varphi_x u_t + \varphi_y v_t + \varphi_{xt} u + \varphi_{yt} v + \varphi_{tt} = 0 .$$

Now we can substitute u_t , v_t into (12c) from (1a), (1b) and u_{xt} , v_{yt} from (12a), (12b) to get a new equation for u , v containing no time derivatives of u , v :

$$(13) \quad \begin{aligned} & uu_{xx} + vu_{xy} + uv_{xy} + vv_{yy} + u_x^2 + 2u_y v_x + v_y^2 + f(u_y - v_x) \\ & + \varphi_x (uu_x + vu_y + \varphi_x - fv) + \varphi_y (uv_x + vv_y + \varphi_y + fu) \\ & - \varphi_{xt} u - \varphi_{yt} v + \varphi_{xx} + \varphi_{yy} - \varphi_{tt} = 0 . \end{aligned}$$

Together with (11) this would then yield a set of compatibility conditions similar to (5). We have however to verify whether (11) and (13) are indeed independent equations. Since (13) is a second-order equation, the standard procedure to test for independence is to introduce new dependent variables in order to convert (11) and (13) into a first-order system.

Letting

$$u_y = q, \quad v_x = r,$$

one such equivalent system for the six unknowns u, v, p, q, r, s would be

$$\begin{aligned}
 u_x &= p, \\
 v_y &= s, \\
 p_x + s_x + \phi_x p + \phi_y r + \phi_{xx} u + \phi_{xy} v &= -\phi_{tx}, \\
 (14) \quad p_y - q_x &= 0, \\
 s_x - r_y &= 0, \\
 u p_x + v p_y + u s_x + v s_y + p^2 + 2qr + r^2 + f(q-r) \\
 &+ \phi_x (up + vq - fv) + \phi_y (ur + vs + fu) - \phi_{xt} u - \phi_{yt} v \\
 &= \phi_{tt} - \phi_{xx} - \phi_{yy} - \phi_x \phi_x - \phi_y \phi_y,
 \end{aligned}$$

where we differentiated (11) with respect to x to obtain the third equation.

A curve is called characteristic for the system (14) if a linear combination of the equations of (14) contains only

interior differential operators for that curve, i.e., directional derivatives along that curve. Define

$$\xi = \psi_x, \quad \eta = \psi_y$$

to be the components of the normal to a characteristic curve

$$\psi(x, y) = \text{const.}$$

of the system (14), and let

$$\tau = u\xi + v\eta .$$

Then the characteristics of the system (14) will be given by

$$\det \begin{pmatrix} \xi & & & & & \\ \eta & & & & & \\ & \xi & & & \xi & \\ & \eta & -\xi & & & \\ & & & -\eta & \xi & \\ & \tau & & & & \tau \end{pmatrix} = 0$$

(e.g., Courant and Hilbert, 1962). Developing this determinant yields

$$(15) \quad \xi^2 \eta (\xi \eta \tau - \xi \eta \tau) \equiv 0 .$$

Therefore, every curve is a characteristic curve. Hence the matrix differential operator on the vector (u, v, p, q, r, s) defined by (14) is an interior operator on every possible curve of the x, y -plane, in other words the equations in system (14), and therefore equations (11) and (13) cannot be independent.

Comparing the manipulations which led to equation (13) to those which led to equation (4), we could have anticipated this result. Indeed, in (12c) we had to use both (1a), (1b) in order to eliminate u_t , v_t and (12a), (12b) in order to eliminate u_{xt} , v_{yt} . But (12a), (12b) were derived from (1a), (1b) by differentiation and thus we can say that we had to use (1a), (1b) twice in order to obtain (13). On the other hand only one use of (2a), (2b) was necessary to obtain (4). The obvious analogy with a similar situation in handling linear algebraic systems is the intuitive content of our result (15).

To understand the deeper reason for the failure of the procedure which was successful in the case of system (2) we will have to consider the general case of the primitive equations.

3. Compatibility conditions for the primitive equations

For the sake of simplicity we shall discuss the primitive equations in x, y, p -coordinates, but the same results can easily be shown to hold in x, y, z - and x, y, σ -coordinates, as they obviously should. We take the equations in the form

$$(16a) \quad u_x + v_y + \omega_p = 0 ,$$

$$(16b) \quad u_t + uu_x + vu_y + \omega u_p - fv = -\phi_x ,$$

$$(16c) \quad v_t + uv_x + vv_y + \omega v_p + fu = -\phi_y ,$$

$$(16d) \quad \theta_t + u\theta_x + v\theta_y + \omega\theta_p = 0 ,$$

where ϕ is the geopotential of an isobaric surface, θ is potential temperature and $\omega = dp/dt$ is the vertical velocity in this coordinate system.

Note that whereas this system is totally hyperbolic, i.e., it has four families of characteristics (three of which coincide, v. discussion of characteristics and their determination infra) the three-dimensional hyperplane $t = 0$ in the four-space (x, y, p, t) is not free.* The plane $t = 0$ is however free for the primitive equations in x, y, z - and x, y, σ -coordinates (where a time derivative appears also in the continuity equation). Hence a Goursat problem (John, 1971, p. 191) with part of the data prescribed on

* The Cauchy problem for a first-order hyperbolic system consists in finding a solution of the system with values of the unknown variables prescribed on some hypersurface. A surface is then called free if the Cauchy problem with data on it has a solution "in the small" (i.e., at least in some neighborhood of the surface), characteristic otherwise (John, 1971, p. 54 ff.). In particular a Cauchy problem with data on $t = 0$ is called a pure initial-value problem.

$t = 0$, part on $p = 0$ is adequate for our system, rather than a pure initial-value problem.

Again we assume that ϕ is known, together with all its necessary derivatives, from the hydrostatic equation

$$\phi_p = -RT/p ,$$

where R is the gas constant. We note in passing that this discussion will be valid no matter what known right-hand sides equations (16) might have.

We want to find a set of equations in u , v , ω , ϕ , θ containing no time derivatives of u , v , ω , or, as we said before, a set of compatibility conditions the wind field has to satisfy instantaneously. Since no equation for ω_t exists in our system, it is natural to use (16d) to eliminate ω ,

$$(17) \quad \omega = -\frac{1}{\theta_p} (\theta_x u + \theta_y v + \theta_t) .$$

Differentiating this with respect to p and substituting into (16a) we obtain

$$(18) \quad u_x + \theta^{(1)} u_p + v_y + \theta^{(2)} v_p + \theta^{(3)} u + \theta^{(4)} v = -\theta^{(5)} ,$$

where

$$\theta^{(1)} = -\theta_x/\theta_p , \quad \theta^{(2)} = -\theta_y/\theta_p ,$$

$$\theta^{(3)} = \theta_x \theta_{pp}/\theta_p^2 - \theta_{xp}/\theta_p , \quad \theta^{(4)} = \theta_y \theta_{pp}/\theta_p^2 - \theta_{yp}/\theta_p ,$$

$$\theta^{(5)} = \theta_t \theta_{pp}/\theta_p^2 - \theta_{tp}/\theta_p$$

are known according to our standing assumption, since T and p are.

To eliminate the time derivatives of u , v between (18) and (16b), (16c), we differentiate (18) with respect to t , and (16b), (16c) with respect to p and with respect to x , y , to yield

$$(18') \quad u_{xt} + \theta^{(1)} u_{pt} + v_{yt} + \theta^{(2)} v_{pt} + \theta^{(3)} u_t + \theta^{(4)} v_t \\ + \theta_t^{(1)} u_p + \theta_t^{(2)} v_p + \theta_t^{(3)} u + \theta_t^{(4)} v = -\theta_t^{(5)},$$

$$(19a) \quad u_{xt} + uu_{xx} + vu_{xy} + \omega u_{xp} + u_x^2 + u_y v_x + u_p \omega_x - fv_x = -\phi_{xx},$$

$$(19b) \quad u_{pt} + uu_{xp} + vu_{yp} + \omega u_{pp} + u_x u_p + u_y v_p + u_p \omega_p - fv_p = -\phi_{xp},$$

$$(19c) \quad v_{yt} + uv_{xy} + vv_{yy} + \omega v_{yp} + u_y v_x + v_y^2 + v_p \omega_y + fu_y = -\phi_{yy},$$

$$(19d) \quad v_{pt} + uv_{xp} + vv_{yp} + \omega v_{pp} + u_p v_x + v_y v_p + v_p \omega_p + fu_p = -\phi_{yp}.$$

Substituting u_{xt} , u_{pt} , v_{yt} , v_{pt} from (19) and u_t , v_t from (16b), (16c) into (18') we obtain

$$(20) \quad uu_{xx} + vu_{xy} + \omega u_{xp} + uv_{xy} + vv_{yy} + \omega v_{yp} \\ + \theta^{(1)}(uu_{xp} + vu_{yp} + \omega u_{pp}) + \theta^{(2)}(uv_{xp} + vv_{yp} + \omega v_{pp}) \\ + u_x^2 + u_y v_x + u_p \omega_x + u_y v_x + v_y^2 + v_p \omega_y \\ + \theta^{(1)}(u_x u_p + u_y v_p + u_p \omega_p) + \theta^{(2)}(u_p v_x + v_y v_p + v_p \omega_p) \\ + \theta^{(3)}(uu_x + vu_y + \omega u_p) + \theta^{(4)}(uv_x + vv_y + \omega v_p) \\ + f(u_y - v_x) + f(\theta^{(2)} u_p - \theta^{(1)} v_p) - \theta_t^{(1)} u_p - \theta_t^{(2)} v_p \\ + f(\theta^{(4)} u - \theta^{(3)} v) - \theta_t^{(3)} u - \theta_t^{(4)} v \\ = -\phi_{xx} - \phi_{yy} - \theta^{(1)} \phi_{xp} - \theta^{(2)} \phi_{yp} - \theta^{(3)} \phi_x - \theta^{(4)} \phi_y + \theta_t^{(5)}.$$

Here again ω , ω_x , ω_y , ω_p can be eliminated with the aid of (17), and only first-order space derivatives of u , v appear in the process. Thus we can view (20) as a second-order equation in u , v .

To test the independence of (18) and (20) it is convenient to introduce

$$u_y = m, \quad u_p = r,$$

$$v_x = n, \quad v_p = t,$$

and differentiate (18) with respect to p . Then we obtain the following first-order system

$$u_x = q,$$

$$v_y = s,$$

$$m_x - q_y = 0,$$

$$n_y - s_x = 0,$$

$$(21) \quad q_p - r_x = 0,$$

$$s_p - t_y = 0,$$

$$uq_x + vq_y + (\omega + \theta^{(1)} u)q_p + \theta^{(1)}(vr_y + \omega r_p)$$

$$+ us_x + vs_y + (\omega + \theta^{(2)} v)s_p + \theta^{(2)}(ut_x + \omega t_p) + \dots = 0,$$

$$r_x + \theta^{(1)}r_p + t_y + \theta^{(2)}t_p + \dots = 0,$$

where the dots stand for terms which contain only the unknowns

$U = (u, v, m, n, q, r, s, t)$ in non-differentiated form or not at all.

Thus system (21) can be written as

$$A_1 U_x + A_2 U_y + A_3 U_p + B = 0 ,$$

where A_1, A_2, A_3 are matrices depending on U , but not on its derivatives and B is a column vector with the same property.

Let a surface in x, y, p -space be given by an equation

$$\psi(x, y, p) = \text{const.}$$

and denote by

$$\xi = \psi_x, \quad \eta = \psi_y, \quad \zeta = \psi_p$$

the components of its normal. Consider the matrix

$$A = A_1 \xi + A_2 \eta + A_3 \zeta =$$

$$\begin{pmatrix} \xi & & & & & & & \\ & \eta & & & & & & \\ & & \xi & -\eta & & & & \\ & & & & & & -\xi & \\ & & & & \zeta & -\xi & & \\ & & & & & & & \\ \tau + (\omega + \theta^{(1)} u) \zeta & \theta^{(1)} (v \eta + \omega \xi) & \tau + (\omega + \theta^{(2)} v) \zeta & \theta^{(2)} (u \xi + \omega \zeta) & & & & \\ & \xi + \theta^{(1)} \zeta & & & & \eta + \theta^{(2)} \zeta & & \\ & & & & & & \zeta & -\eta \end{pmatrix},$$

where we used a slight reordering of the equations of (21) (the equation which was considered the fifth one in the first writing being now last) and where $\tau = u\xi + v\eta$.

By the theory of characteristics the quasi-linear system (21) is elliptic at a point (x, y, p) for a certain solution $U = U(x, y, p)$ if the equation

$$(22) \quad \det A(x, y, p; U; \xi, \eta, \zeta) = 0$$

has no real solution (ξ, η, ζ) there and is totally hyperbolic if eight distinct real solutions $(\xi, \eta, \zeta)_i$, $i = 1, 2, \dots, 8$, exist. Any such "solution" of (22) is a first-order partial differential equation for a function ψ . The solution of this PDE yields a family of characteristic surfaces

$$\psi(x, y, p) = \text{const.}$$

of the system (21). The equations of (21) cannot be independent of each other if (22) is identically satisfied, irrespective of the values of the variables on which A depends, as was already explained for the shallow-fluid equations. Indeed, for (22) to be identically satisfied, it is necessary that the matrix differential operator

$$C = A_1 \frac{\partial}{\partial x} + A_2 \frac{\partial}{\partial y} + A_3 \frac{\partial}{\partial p}$$

be an interior operator at any point of any surface in space for any solution U of (21). In other words, a solution \tilde{U} of some of the equations in system (21) must satisfy the whole system

$$C(\tilde{U})\tilde{U} + R(\tilde{U}) = 0 .$$

Developing the determinant in (22) by minors yields

$$\begin{aligned}
\det A &= \xi^2 \eta^2 \left[\zeta \{ -\theta^{(1)} \zeta (v\eta + \omega\xi) (\eta + \theta^{(2)} \zeta) + \eta [\tau + (\omega + \theta^{(2)} v) \zeta] (\xi + \theta^{(1)} \zeta) \right. \\
&\quad \left. + \theta^{(2)} \zeta (u\xi + \omega\xi) (\xi + \theta^{(1)} \zeta) \} - \xi [\tau + (\omega + \theta^{(1)} u) \zeta] (\eta + \theta^{(2)} \zeta) \zeta \right] \\
&= \xi^2 \eta^2 \zeta \{ [\theta^{(2)} (u\xi^2 + \omega\xi\xi + u\theta^{(1)} \xi\xi) - \theta^{(1)} (v\eta^2 + \omega\eta\xi + v\theta^{(2)} \eta\xi)] \zeta \\
&\quad + (\theta^{(2)} v - \theta^{(1)} u) \xi \eta \zeta + \theta^{(1)} [u\xi + v\eta + (\omega + \theta^{(2)} v) \zeta] \eta \zeta \\
&\quad - \theta^{(2)} [u\xi + v\eta + (\omega + \theta^{(1)} u) \zeta] \xi \zeta \} \equiv 0 .
\end{aligned}$$

Thus the equations of system (21) are not independent. Hence the same holds for equations (18) and (20), to which system (21) is equivalent, given suitable side conditions.

We remark moreover that neither is a full set of compatibility conditions possible for the complete Euler equations, in which the equation for the vertical velocity cannot be reduced to the hydrostatic equation. Indeed, let us assume that somehow all variables of state, i.e., T , p , θ are known together with the necessary derivatives. Even then, differentiating the energy equation equivalent in x, y, z -coordinates to (16d) with respect to t and substituting u_t , v_t , w_t from the three momentum equations into the equation so obtained will lead to one compatibility condition. The continuity equation will then give just one additional condition, and it is verified by inspection that no other independent equation in which time derivatives of (u, v, w) do not appear can be derived. Thus we have again an underdetermined set of two equations for the instantaneous determination of the three unknowns u , v , w .

In this context it is worthwhile to note that in Mesinger's (1972) iterative updating experiments the use of the first time derivative of the state variable brought a significant improvement in the adjustment of the wind field. The use of the second time derivative however did not. This reflects the fact, which should be clear in the light of our discussion, that no additional independent compatibility conditions can be obtained by further differentiation with respect to time of the energy or continuity equations.

Obviously, to consider the Navier-Stokes equations, or other equations with forcing and dissipation depending on the wind field, would make matters only worse. This seems to exhaust the possibilities.

We are therefore forced to conclude that there is no adequate generalization of exact geostrophic balance to atmospheric phenomena governed by the primitive Euler or Navier-Stokes equations, at least not in the sense that the wind field is entirely and instantaneously determined by the thermodynamic state of the atmosphere and its past history as reflected in the time derivatives of the state variables.

4. Discussion and conclusions

The analysis above should make it clear why the mathematical argument mentioned in the Introduction and used by some authors to justify the updating approach to the initialization problem may not be appropriate. Not only are the systems of PDEs under consideration non-linear, which in general precludes the elimination of all but one variable; the even more pertinent difficulty is that such elimination does not lead to a well-posed problem to determine the wind field from temperature data.

On the other hand, our analysis proves rather conclusively that for an atmosphere governed by the primitive equations even a generalized geostrophic balance, in the sense of compatibility conditions as previously defined, can be only an approximate, rather than an exact mathematical property.

We think that this discussion at least partly explains the difficulties encountered by the updating and four-dimensional data assimilation approaches to the initialization problem.

Specifically these difficulties are: (i) that the errors in the initial data cannot be reduced to zero or to negligible values, (ii) that in forward updating the asymptotic value of the error is reached after times comparable to the predictability time, and (iii) that use of data extraneous to the numerical model used seems to lead to catastrophic results. The partial success of these approaches has probably to do with a rather complicated mechanism, involving the forcing and dissipation in the system of equations used, which we cannot discuss here and which is currently under investigation.

Our analysis on the other hand has shown that instantaneous compatibility conditions for the velocity field can be derived, but that their number (one for the non-linear shallow-fluid equations, two for the primitive equations) is in general one less than the number required to determine the velocity field from the thermodynamic fields and their history. This seems to justify the contention that it is necessary, in order to solve the initialization problem, to be able to measure at least one scalar parameter determining the wind field. This scalar could be either the (horizontal) wind direction or the wind speed.

The wind direction could possibly be determined by cloud tracking techniques which are not at present sufficiently accurate in order to determine also the speed. The speed on the other hand could perhaps be determined with sufficient accuracy (concerning the present GARP requirements see for instance Kasahara, 1972) by other methods which are still easier to implement than those requiring a full and accurate measurement of two or three velocity components separately.

Our analysis may be extended to study four-dimensional data assimilation in the following way. Assume that one would wish to solve a Cauchy problem for a hyperbolic system of equations of large-scale dynamic meteorology (the shallow-fluid equations or the primitive equations) different from the pure initial-value problem. This would correspond to data being available on some arbitrary hypersurface S in (x, y, z, t) four-space, rather than on the hyperplane $t = 0$, e.g., on a surface

spanned by the trajectories of satellites. If the data were complete and the given surface S were free for the system of equations used, as well as being a space-like surface (John, 1971, p. 113 ff., Courant and Hilbert, 1962, p. 589 ff.), then the Cauchy problem with respect to S would be well posed.

If the data, however, are not complete, as is to be expected, although S is free and space-like, we have again with respect to S an "initialization problem" or "incomplete Cauchy problem". This can now be handled either by "updating techniques", as is currently done in various numerical experiments, or else by the analytical methods proposed above. Our analysis of the initial-value problem seems to suggest that a theoretical investigation of four-dimensional data assimilation in the sense outlined here might help direct the numerical experiments in such respects as optimal number, nature and location of data, or assessment of the completeness and relative value of data available under present and future data gathering programs.

This subject, as well as the numerical implementation and practical evaluation of the theoretical methods proposed here for the initial-value problem, are intended to be covered in future work by the author.

References

Charney, J. G., 1955: The use of the primitive equations of motion in numerical prediction, Tellus, VII, 22-26.

Charney, J. G., R. G. Fleagle, V. E. Lally, H. Riehl and D. O. Wark, 1966: The feasibility of a global observation and analysis experiment, Publication 1290, Natl. Acad. Sci./Natl. Res. Council, Washington, D.C.

Charney, J., M. Halem and R. Jastrow, 1969: Use of incomplete historical data to infer the present state of the atmosphere, J. Atmos. Sci., 26, 1160-1163.

Courant, R. and D. Hilbert, 1962: Methods of Mathematical Physics, Vol. II, John Wiley and Sons, New York.

Haltiner, G. J., 1971: Numerical Weather Prediction, John Wiley and Sons, New York.

John, F., 1971: Partial Differential Equations, Springer, New York.

Kasahara, A., 1972: Simulation experiments for meteorological observing systems for GARP, Bull. Amer. Meteor. Soc., 53, 252-264.

Mesinger, F., 1972: Computation of the wind by forced adjustment, J. Appl. Meteor., 11, 60-71.

Morel, P., G. Lefevre and G. Rabreau, 1971: On initialization and non-synoptic data assimilation, Tellus, XXIII, 197-206.

Nitta, T. and J. B. Hovemale, 1969: A technique of objective analysis and initialization for the primitive forecast equations, Mon. Wea. Rev., 97, 652-658.

Williamson, D. and A. Kasahara, 1971: Adaptation of meteorological variables forced by updating, J. Atmos. Sci., 28, 1313-1324.

Williamson, D. L. and R. E. Dickinson, 1972: Periodic updating of meteorological variables, J. Atmos. Sci., 29, 190-193.